

Agent Evaluation Metrics — One-Page Cheat Sheet

Retrieval & Discovery

Metric	What it Measures
Recall / Recall@K	Coverage of all relevant items
Precision / Precision@K	Relevance of retrieved items
Hit Rate	Whether any relevant item was found

Ranking Quality

Metric	What it Measures
MRR	How early the first relevant result appears
NDCG	Overall ranking quality with graded relevance
MAP	Average precision across ranked results

Reasoning & Task Success

Metric	What it Measures
Exact Match / F1	Correctness of final answer
Task Success Rate	End-to-end completion
Utility Score	Practical usefulness of output

Faithfulness & Efficiency

Metric	What it Measures
Groundedness / Citation Accuracy	Claims supported by evidence
Hallucination Rate	Unsupported statements
Latency / Token Usage	Speed and cost efficiency